Visual detection of coarticulatory anticipation or... guessing what has not yet been written

Sonia KANDEL*, Louis-Jean BOË* & Jean-Pierre ORLIAGUET**

* Institut de la Communication Parlée, CNRS, Univ. Stendhal/INPG-ENSERG
Domaine Universitaire 1180 Av. Centrale - BP 25 - 38040 Grenoble Cedex 9 - France
Tel. (33) 76 82 43 37 - Fax (33) 76 82 43 35
e-mail: kandel@icp.grenet.fr - boe@icp.grenet.fr

** Laboratoire de Psychologie Expérimentale, CNRS, Univ. Mendès France
BP 47 X - 38040 Grenoble Cedex- France

Abstract

Emerging technology in communication systems requires the consideration of specific psycholinguistic phenomena for the conception of optimally efficient devices. The reported study focuses on a phenomenon observed in speech and cursive handwriting, namely, the visual detection of coarticulatory anticipation, and examines, in particular, the contribution of visible motor information in linguistic message decoding. Visual dynamic and static digram and word stimuli –prepared by reproducing identical spatial templates, and on the basis of the presence of coarticulatory anticipation in the handwriting movements– were administered without providing spatial cues on the identity of the subsequent letter. Results indicate that when kinematic information was provided in the dynamic visual stimulus, subjects were able to predict with fair accuracy the identity of the following letter. When the only information supplied by the stimulus was the static trace of the production, subjects could not predict the identity of the following one. Therefore, the visual system detects temporal differences that translate anticipatory gestures in the production movements.

Introduction

Communication systems and human factors

The development of communication systems, does not only require high technology material. In the conception of optimally efficient devices, specific psycholinguistic phenomena must also be taken into consideration. As Benoît (1992) has pointed out, the quality of emerging technology in audiovisual systems such as animated talking faces, the movie industry, Text-to-Audio-Visual speech synthesizers (TxAVS), multimodal man-machine dialogue, visiophone, videoconference, is conditioned by the manner in which human beings produce and perceive bimodal speech. Humans produce speech by generating movements of several articulators, of which only some are visible (lips, jaw). Some speech movements are only visible, some are only audible, and some are audible, as well as visible (Abry & Schwartz in Cathiard, 1988/89). In addition, continuous speech is not continuously audible. Speech is naturally made of significant parts of silence in which the speaker makes visible labial gestures that anticipate the following sequence. Humans take advantage of the intrinsic bimodality of the speech communication system: information is simultaneously transmitted through two channels –the acoustic and the optic flows– and the outputs are integrated by the perceiver.

0-7803-1363-1/93 $3.00

Therefore, the way in which auditory and visual information will be treated, would depend on the transmission quality of the audiovisual system to be developed. Findings in bimodal speech research show that: a) visual information considerably enhances intelligibility of audiovisual speech, especially when acoustic information is degraded (Sumby & Pollack); b) this speech-reading benefit generally holds even when the channels are slightly desynchronized (Campbell & Dodd, 1980); c) vision can sometimes bias message comprehension in conflicting auditory and visual presentations (McGurk & McDonald, 1976); d) due to anticipatory coarticulation, the eye often receives information before the ear and seems to take advantage of it (Cathiard & Lallouache, 1992). The reported study focuses on the latter finding, namely, the visual detection of coarticulatory anticipation.

Visual detection of coarticulatory anticipation in speech

Coarticulation is inherent to transitions between speech sounds. At an articulatory level, Benguérel & Cowan (1974) have shown that labial protrusion for the rounding feature of a French vowel [y] may anticipate, throughout a consonant string, the onset of its voiced output. The perceptual efficiency of this anticipation phenomenon has been demonstrated by Cathiard & Lallouache (1992). The latter provide evidence that motor anticipation in the lip rounding gesture is detected by the visual system and that this information is exploited in phonetic decoding, in the absence of acoustic information: during an acoustic pause, subjects could visually identify the protrusion gesture of a following [y] up to 210 ms before the onset of the sound. Thus, visual detection of coarticulatory anticipation enables subjects to perceive what they have not yet heard. The visual system recovers spatial information (labial geometry), but also kinematic information (labial anticipatory gestures): a difference of 30 msec is observed between dynamic and static presentations of the stimuli. Cathiard & Lallouache (1992) point out, however, that although this difference is significant, the temporal gain is relatively small in comparison to the temporal amplitude of the labial gesture. Therefore, the dynamic presentation of the stimulus does not seem to be so crucial for the visual detection of coarticulatory anticipation in bimodal speech perception. Finally, they argue that this does not imply that kinematic information was not perceived in the static presentations, since Freyd (1983) has shown that movements can also be recovered from static stimuli.

Visual detection of coarticulatory anticipation in handwriting

In the reported experiment, we aimed on one hand, at showing the visual detection of coarticulatory anticipation phenomenon in cursive handwriting and, on the other, at examining the effect of static and dynamic presentations on the identification of subsequent information.

Studies on the production of handwriting movements show that allographic effects can influence kinematic patterns. Orliaguet & Boë (1990), have shown that temporal patterns observed in letter writing may be subject to variations according to the graphemic context in which they are produced. In reproducing a spatially defined template of the same letter *l*, the timing of the down-stroke is function of the topological constraints of the following letter: changes in size (*ll vs. le*) as well as in size and rotation direction (*ll vs. ln*), are translated by differences in temporal patterns. This means that the motor system *anticipates* the topological changes of the subsequent letter, modifying the timing of the current production. As for the elements of speech production, kinematic patterns in the handwriting of an identical letter vary according to specific adjustments that are due to contextual constraints.

On this basis, it was aimed on one hand, at providing evidence for the visual detection of the coarticulatory anticipation phenomenon observed in the time course of cursive handwriting by Orliaguet & Boë (1990) and, on the other, at examining the effect of static and dynamic presentations on the identification of subsequent information, in a situation in which the geometric differences of the stimuli are neutralized. Thus, dynamic and static visual digram and word stimuli were presented, to determine whether the relevant information in this type of linguistic identification task, is the kinematics of the production movement.

Method

Subjects

16 French-speaking male and female right-handed subjects (since items were produced by a right-handed person) volunteered to participate in the experiment. All of them were students of several domains, between 20 and 30 years old, without any particular knowledge concerning movement control nor visual perception.

Apparatus, procedure and task

On the basis of Orliaguet & Boë (1990)'s results for handwriting, we used digrams (*ll* - *le* - *ln*) and words (*fille* - *filet* - *filme*) containing specific spatial constraints that the motor system anticipates by velocity changes. An identical spatial template of an *l* was used to reproduce spatially defined *ll le lm* and *fille filet filme* sequences on a Numonics 2200 - 0.30.E graphic tablet. The tablet yielded *xy* coordinates recorded at a frequency of 100 Hz.

In each of the recorded handwritten sequences, the anticipation of the topological constraints of the subsequent letter were reflected by temporal differences in the down-stroke of the *l*. Then, the productions were cut at the lowest point of the down-stroke of the *l* such that no spatial information on the identity of the following letter should be provided by the stimulus (see figure 1). The experiment was run in a HyperCard 2.1 environment, and a PASCAL extension (XFCN) permitted us to reproduce the exact dimensions, velocity and the temporal course of the original productions on the screen of a IIFX Macintosh, where the experiment was carried out.

Before the presentation of each sequence, the subject was told that he/she would see an *l* or a *fil* on the computer screen. The task consisted in *guessing* whether the presented *l* corresponded to the production of *ll*, *le* or *lm* (for digrams) and whether the presented *fil* corresponded to the production of *fille*, *filet* or *filme* (for words). Once the subject gave the answer (by choosing with the mouse one of the three possibilities appearing on the screen), the following item was presented, without knowledge of results.

The *dynamic* condition consisted of a visual presentation of the original productions, progressively traced on the screen with their exact dimensions, velocity and the temporal course. They disappeared from the screen as soon as the lowest point of the down-stroke of the *l* was reached. In the *static* condition, the same productions were presented, but an entire item at once: the entire *l* (for digrams) and the entire *fil* (for words) were "flashed" on the screen. The presentation times of both conditions was the same since it corresponded to the production time of the item. Order of condition presentation was counterbalanced, whereas the order of presentation of the stimuli was randomized. All experimental conditions were presented to the subjects. Each condition consisted of 36 random presentations, 12 for each item: *ll le lm fille filet filme* (each subject therefore saw 144 presentations).

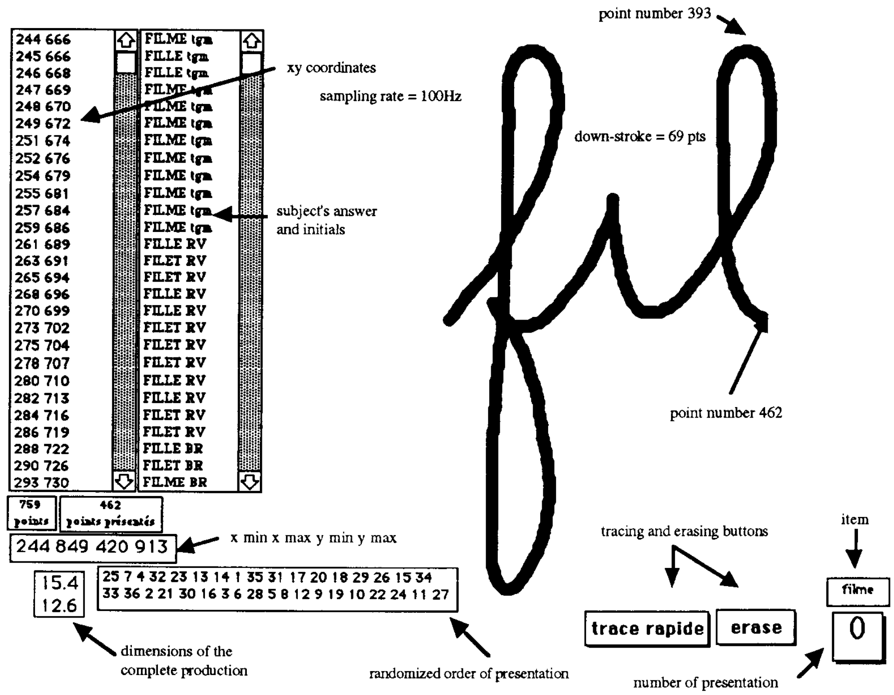

*Figure 1. Description of the **fil** corresponding to **filme** . The subject only saw **fil** appearing on the screen and then had to "guess" whether it corresponded to the production of **filme**, **filet** or **fille**.*

Results

Mean percentages of correct response as a function of the different experimental conditions are plotted in figure 2. An ANOVA with presentation mode (dynamic - static), spatial change following the *l* (*l* - *e* - *m*), and lexical value (digrams - words) as within-subject factors, was carried out. The dependent measure concerned the percentage of correct responses. Data analysis reveals a significant presentation mode main effect ($F(1, 15) = 47.82$, $p < .001$). Differences between the dynamic and the static conditions are significant for all stimuli (t-tests are significant at $p < .01$). A significant interaction between the presentation mode and the spatial change condition ($F(2, 30) = 5.324$, $p < .01$) indicates that differences between dynamic and static conditions vary as a function of the spatial change of the letter following the *l*. In addition, global results for the static condition do not differ from chance (33.3%): all t-tests are non-significant except for stimulus *ll* ($p < .01$), which is highly below chance. Correct identification in the dynamic condition is highly above chance from the onset of the test (t-tests are all significant at $p < .01$).

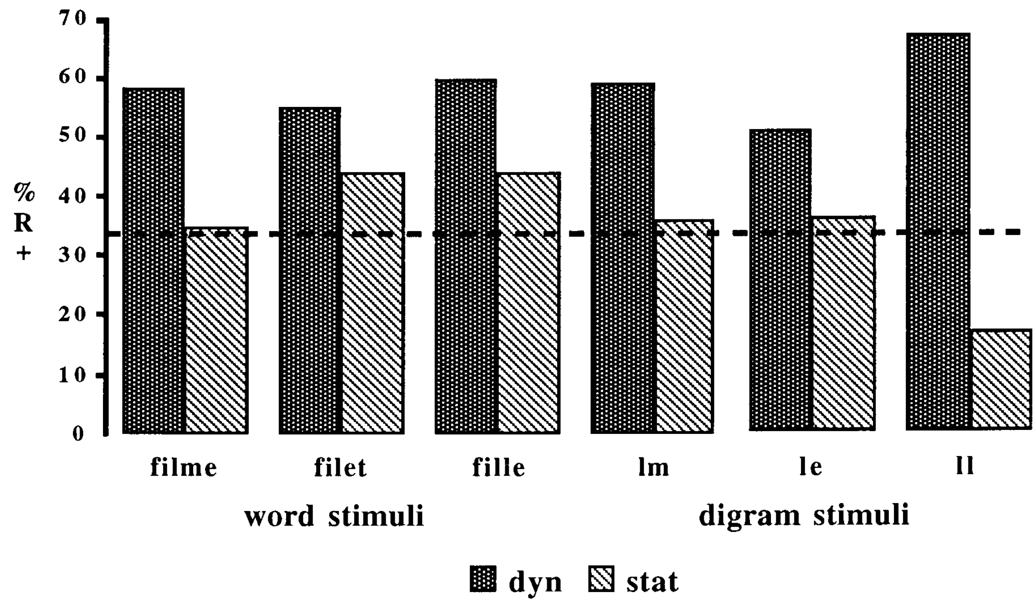

Figure 2. Mean percentage of correct response as a function of the presentation mode (dynamic - static), spatial change following the l (l - e - m) and lexical value (words - digrams). Chance corresponds to 33.3%.

Results indicate that the dynamic condition enhances the prediction of the identity of the letter following the *l* in comparison to the static condition. Data also suggest that visual detection of coarticulatory anticipation in handwriting is not a learned skill, since the percentage of correct identification for the first four digram and word items is highly above chance (mean 54%). Therefore, results for the dynamic condition provide evidence for a visual detection of anticipatory velocity changes observed in the time course of digram and word productions.

Discussion

The results of this experiment show that the visual system detects the coarticulatory anticipation phenomenon observed in handwriting production movements, only in the dynamic presentation. Results indicate that in the dynamic visual stimulus, subjects were able to predict with fair accuracy the identity of the following letter; i.e., they were able to exploit the kinematic anticipatory gestures of the production. It thus seems that the information on the identity of the following letter can be provided by the kinematic differences in the anticipatory production movements. These temporal differences translate the topological constraints –such as size and/or rotation direction changes– of the shapes of the following letters. This performance is observed in the handwriting of digrams as well as words.

Research on the perception of static handwritten traces suggests that recognition of handwriting involves feature geometric analysis, and also, knowledge on the actual movements used to trace the letters (Babcock & Freyd, 1988). Indeed, results presented by these authors indicate that perceivers spontaneously infer the underlying dynamic pattern of motor movements used to produce handwritten characters by applying their own knowledge of the drawing method (i.e., the general rules of letter production) to its static trace. In recognizing the presented static traces, perceivers seemed to capitalize on shared knowledge –both by producers and perceivers– on the dynamics of specific production rules. Indeed, the stimuli

used by Babcock & Freyd (1988) consisted in spontaneous and complete handwritten static traces; i.e., information on the production movement could be recovered from the continuous trace. Conversely, the static stimuli used in the experiment reported in this paper, were prepared on the basis of a spatially defined template and were cut so that no spatial cues on the identity of the letter following the *l* were provided. That is, the shape of our static stimuli did not contain information on the temporal differences observed in the production movements of the sequences. Thus, in the identification task in this experiment, subjects could not capitalize on implicit procedural knowledge from static traces, since kinematic differences could only be perceived in the dynamic presentation.

Nevertheless, implicit procedural knowledge could be involved in the visual detection of coarticulatory anticipation, in the dynamic condition. Results indicate that what is anticipated in production, can also be anticipated perceptually: the visual system exploits kinematic information, that enable the subject to predict –or *anticipate–* subsequent spatial constraints. This competence necessarily requires some kind of procedural knowledge on the velocity modulations occurring in bidimensional manual movements. Viviani & Stucchi (1992) provide evidence in support of the idea that procedural knowledge can influence the visual perception of dynamic bidimensional human movements. These authors have shown, that motor variables affect visual perception to such an extent, that humans can be subject to geometric and kinematic illusions if motor information is in disagreement with implicit procedural knowledge. Their data show that the visual perception of dynamic stimuli is distorted when the presented movements do not adjust to the characteristics of biological motion (especially, for figural movements like drawing and scribbling). It seems then, that visual perception of bodily motion relies on a specific type of motor information; i.e., dynamic information that respect a biological model of physical constraints.

According to Viviani & Stucchi (1992), an internal, standard movement-control model –which determines movement dynamics according to biological constraints– enables humans to regulate and correctly perceive bodily motion. The motor system interacts with the visual system in the perception of dynamic stimuli. We may thus hypothesize that correct identification of a subsequent letter, on the basis of kinematic information, could be due to the recovery of procedural motor representations. The stimuli dynamic in our experiment, followed the physical constraints of biological motion, since they reproduced the exact dimensions and temporal course of the productions. At present, an experiment is being carried out, in which handwritten kinetic stimuli do not follow the physical constraints of biological motion. If future results indicate that perceptual anticipation is only possible when the stimuli respect the constraints of biological motion, the hypothesis of a recovery of procedural motor representations in perceptual processes will become more plausible.

Conclusion

By showing that one can guess what your hand has not yet written –on the sole basis of the kinematics of the manual gesture– the reported study provides evidence that the motor system may contribute to the visual perception of dynamic *biological* information. The information exploited by the visual system may influence linguistic decoding, by predicting certain characteristics of subsequent sensory information. The visual detection of coarticulatory anticipation in speech could enable the identification of a specific feature, like the one corresponding to lip rounding, in an on-line process of selection of possible word candidates in the mental lexicon. Subsequent acoustic information could then be used as soon as it becomes available, to converge towards a single candidate. Therefore, research on communication systems that aim at conceiving optimally efficient devices, should consider the importance of these perceptual anticipatory phenomena in processes of linguistic decoding.

Aknowledgements

We would like to thank Christian Abry for his very helpful scientific support and Rudolph Sock for suggestions in the manuscript.

References

BABCOCK M.K. & FREYD J.J. (1988). The perception of dynamic information in static handwritten forms. *Amercan Journal of Psychology*, 101, pp. 11-130.

BENGUEREL A.P. & COWAN H.A. (1974). Coarticulation of upper lip protrusion in French. *Phonetica*, 30, 41-55.

BENOÎT C. (1992). The intrinsic bimodality of speech communication and the synthesis of talking faces. *Journal on Communications*, vol. XLIII, pp. 32-40.

CAMPBELL R. & DODD B. (1980). Hearing by eye. Quarterly Journal of Experimental Psychology, 21, 85-99.

CATHIARD M.A. (1988/89). La perception visuelle de la parole: Aperçu de l'état des connaissances. *Bulletin de l'Institut de Phonétique de Grenoble*, 17-18, 109-193.

CATHIARD M.A. & LALLOUACHE M.T. (1992). L'apport de la cinématique dans la perception visuelle de l'anticipation et de la rétention labiales. *Actes des 19e Journées d'Etudes sur la Parole*, Bruxelles, pp. 25-30.

FREYD J.J. (1983). The mental representation of movement when static stimuli are viewed. Perception & Psychophysics, 33(6), 575-581.

McGURK H. & McDONALD (1976). Hearing lips and voices. *Nature*, 264, 746-748.

ORLIAGUET J.-P. & BOË L.-J. L.-J. (1990). Régulation temporelle des mouvements d'écriture en fonction des contraintes spatiales. In V. Nougier & J.P. Bianchi (Eds.), *Pratiques Sportives et Modélisation du Geste*, pp. 163-177.

PERKELL J.S. & MATTHIES M.L. (1992). Temporal measures of anticipatory labial coarticulation for the vowel [u]: within and cross subject variability. *Journal of the Acoustical Society of America*, 91(5), pp. 2911-2925.

SUMBY W.H. & POLLACK I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, 26, 212-215.

THOMASSEN A.J. & SCHOMAKER L.R. (1986). Between-letter context effects in handwriting trajectories in *Graphonomics: Contemporary Research in Handwriting*, H.S. Kao, G.P. Van Galen, R. Hoosain (Eds.), Amsterdam : Elsevier Science Publishers B.V.

VIVIANI P. & STUCCHI N. (1992). Biological movements look uniform: Evidence of motor-perceptual interactions. *Journal of Exper. Psychol. Human Perc. and Perf.* 18, 3 pp. 603-623.